L1 Sign Language Tests and Assessment Procedures and Evaluation

Tobias Haug, Wolfgang Mann, Joanna Hoskin, & Hilary Dumbrill

There is a growing interest in sign language assessments documented by the increase in the number of tests developed over the last decade and also by the literature that is now available on this topic. In this chapter we give an overview of tests for assessing deaf children's signed language skills and discuss challenges related to developing these test as well as to evaluating them in terms of reliability and validity. In addition, we share some experiences on how sign language assessments are used by practitioners. References and links are provided to guide future teachers, researchers, and/or policymakers in their search for further information.

1. Introduction

Sign language tests have been developed for different purposes (Haug, 2005) ranging from monitoring deaf children's sign language development to the assessment of sign language skills in adult learners, who learn a sign language as a second or foreign language, for instance, future interpreters. The term "sign language test for L1 learners" is frequently used to refer to signing deaf children who acquire a sign language at home only in order to differentiate them from adult learners of a sign language, i.e., deaf and hearing adults who learn a sign language as a second or foreign language (e.g., Woll, 2013). However, since also adults can be considered as L1 learners (e.g., codas, deaf adults) as well, we will use the term "young learners" or "early learners" to refer to the linguistically diverse group of deaf and hearing signing children who acquire a sign language from birth or during childhood (up to 6 years old). For the purpose of this chapter, we will only focus on signing children while assessment of adult learners will be

covered in chapter XXX. Readers interested in a more detailed discussion of sign language tests are referred to Enns et al. (2016) or Haug (2005) [1].

2. Theoretical Perspectives

Compared to spoken language assessments the number of sign language tests that are (commercially) available is relatively small. Apart from the fact that sign language research is still a very young field, which only started in the 1960s, this shortage may be a result of specific challenges related to development and evaluation of sign language tests. One such challenge is the incomplete/limited state of research on the structure and acquisition on many sign languages.

For test developers, the difficulty does not arise solely from the lack of documentation of a particular sign language but also when important resources like a reference grammar (Palfreyman, Sagara, & Zeshan, 2015) or a sign language corpus is not available (e.g., Haug, 2017). For instance, if a corpus on the acquisition of sign language is available, this data can be used to inform test development in form of frequency lists of signs as foundation to develop a vocabulary test.

The incomplete state of research is only one possible challenge; another is the small size and heterogeneity of the deaf population. The small number of deaf children makes it difficult to obtain samples that are large enough for norming purposes, i.e., generate average performance scores for different ages. Although this issue might not be a big problem in larger countries such as the United States or some European countries such as Germany, it certainly poses a problem for small countries as well as countries with more than one sign language, e.g., Switzerland with

[1] The website Sign Language Assessment Instruments provides an overview of existing sign language tests https://signlang-assessment.info.

three (Boyes Braem, Haug, & Shores, 2012) and Belgium with two sign languages (Van Herreweghe & Vermeerbergen, 2009).

Another issue is the heterogeneity of group of deaf children in terms of their language acquisition. Deaf children who do not have access to a sign language within the most critical early years of their lives (up to 6 years old; e.g., Mayberry, Lock, & Kazmi, 2002; Newport, 2002) are the main target group for sign language evaluation and intervention (Haug, 2011). The reference group, however, should be deaf and hearing (near native) signing children most of whom come from deaf families. These children serve as models against which the performance of children with late exposure to sign (most deaf children with hearing parents) can be measured to allow for standardization (Herman, 2002; Herman, Holmes, & Woll, 1998).

Different publications (e.g., Mann & Haug, 2014) and guidelines (e.g., Haug et al., 2016) deal with the development and evaluation of sign language tests for deaf children. These publications can serve as a basis for test developers to design tests in a local context.

3. Evaluation of Sign Language Tests

Once a test has been developed, piloted, revised, and a main study with a larger sample has been conducted, the test needs to be evaluated according to certain psychometric properties, i.e., validity and reliability. There are a number of ways how this can be accomplished some of which will be discussed next.

3.1 What are Psychometrics and why are they important?

In order for any (language) test measurements to be trusted, evidence must be presented that the test is valid and reliable. These psychometrics are important as they make it possible to

interpret and generalize the underlying construct that a test measures. The need to report psychometric values is particularly apparent for new assessments that have not been standardized, yet. The following section describes the concept of validity and reliability with concrete examples from existing sign language tests. Since most studies on sign language tests are framed within classical test theory, our examples will be presented within this framework. In addition we will briefly cover more recent approaches to validation.

3.2 Validity

Validity of a test is the understanding that a test truly measures what it is supposed to measure (Kline, 2000). This is notably different from reliability, which determines consistency of a test. There are several types of validity each of which are briefly described below.

3.2.1 Content Validity

Content validity refers to the degree that the instrument covers the content that it is supposed to measure (Bush, 1985). It also refers to the adequacy of the sampling of the content that should be measured. Much of the evidence of content validity collected for existing sign language tests is based on collaborations with native deaf signers or practitioners working with deaf children. For instance, for developing the items of the American Sign Language Vocabulary Test (ASL-VT), Mann and colleagues (2016) worked closely with a panel of deaf and hearing experts. These experts provided feedback related to the multiple-choice format of the test, specifically the target and distractor items. Additional feedback on the quality and clarity of the test images was gathered by a group of hearing undergraduate students. Mann and colleagues also used teacher ratings for the test items to evaluate the type of information (or combination of

types) used by children to acquire these items. Similarly, the developers of the ASL Assessment Instrument (ASLAI; Hoffmeister, 1999) worked closely with a team of native signers with expertise on language development, who advised on the content for each task. In addition, each task was field-tested on a group of ten deaf adults. Only items that showed at least a 90% agreement among the deaf respondents were retained in the item pool (Hoffmeister, 1999). Evidence for content validity of a vocabulary test for German Sign Language (*Deutsche Gebärdensprache*, DGS; Bizer & Karl, 2002a) was provided by using well-established word frequency lists for spoken German as foundation for item selection within the targeted age range (children attending grades 3-5). Finally, Haug (2011) reviewed existing research studies on the linguistic structures of DGS that are represented in a British Sign Language (BSL) test in order to establish content validity during the process of adapting the BSL Receptive Skills Test (Herman et al., 1999) to DGS.

3.2.2 Construct Validity

A second type of validity is construct validity. This type of validity is needed when a test measures a specific attribute or quality, for which there is no operational definition (Cronbach & Meehl, 1955). As a first step, it requires a clear definition of the construct, e.g., intelligence, to be measured (Bechtoldt, 1951). The question which domain or construct should be measured can be determined through a review of the relevant literature, focus groups, and/or interviews (Yaghmaie, 2003). For instance, the underlying construct of the web-based ASL-VT was the assumption that two or more learners may have different knowledge about the same word or sign (Mann et al., 2016). This construct referred to as of construct of strength of form-meaning mappings is illustrated in Table 1. It consists of four levels of mapping each of which

representing one task of the ASL-VT. The levels range from 1, the weakest mapping (meaning recognition) – 4, the strongest mapping (meaning recall).

Table 1: Construct of strength of form-meaning mappings in ASL (adapted from Mann et al., 2016)

Type of mapping		*Task description*
	Strong	
4. Meaning Recall	Strength of mapping ↕	Produce three ASL responses to a sign prompt
3. Form Recall		Produce the target ASL sign for a picture prompt
2. Form Recognition		Match a picture prompt with one of four ASL signs
1. Meaning Recognition		Match a prompt in ASL with one of four pictures
	Weak	

In order to provide a developmental picture of vocabulary growth in ASL, test takers' performances on the different tasks were correlated with age, followed by a comparison of their performances across tasks.

In case of the vocabulary test for DGS (Perlesko: *Prüfverfahren zur Erfassung Lexikalisch-Semantischer Kompetenz*; Bizer & Karl, 2002a) construct validity was established by correlating intra-individual factors with the test results. These factors were, for example, (1) grade attended in school, (2) educational policy of the school (oral, sign, or bilingual), (3)

hearing status of the parents, (4) gender, and (5) chronological age. The results show that the construct "knowledge of receptive vocabulary" was represented by the Perlesko. In comparison, developers of the Assessment for Sign Language for the Netherlands (*Nederlandse Gebarentaal*, NGT) (Hermans, Knoors, & Verhoeven, 2010) carried out three types of correlation analyses to investigate construct validity: correlations between test takers' ages and test performance, between gender and test performance, and between parental hearing status and test performance. Similarly, developers of the ASL-Proficiency Assessment (ASL-PI; Maller, Singleton, Supalla, & Wix, 1999) divided children into three groups, based on their different linguistic experience based on the assumption that they would perform significantly differently on the test. This was confirmed by the results. The ASL-PI is an assessment instrument to measure expressive ASL skills in non-native deaf children.

An alternative approach of presenting evidence in support of construct validity was used for the Language Proficiency Profile (LPP; Bebko & McKinnon, 1993), a test that evaluates children's overall linguistic/communicative, independent of any specific language or modality of expression: Each item was printed separately on a card and presented to three experts, i.e., psycholinguists/language pathologists, who were then asked to sort these items into developmental order within each subscale of the test. Results showed high agreement between raters' ordering and the original ordering (84% across all subscales).

3.2.3 Criterion-related Validity

There are two types of criterion-related validity, concurrent and predictive validity. Concurrent validity is studied when "one test is proposed as a substitute for another, or a test is shown to correlate with some contemporary criterion (e.g., psychiatric diagnosis)" (Cronbach &

Meehl, 1955, p. 282). For example, to collect evidence for concurrent validity, the developers of the ASL-VT compared test takers performance on the four vocabulary tasks with their scores on the ASL Receptive Skills Test (ASL-RST; Enns & Herman, 2011). Although both repeated sets of bivariate correlations and partial correlations with age partialed-out did not show any differences a closer inspection of the distribution of scores on the vocabulary tasks and ASL-RST revealed certain similarities between the ASL-VT and ASL-RST in their profiles across age bands (see Mann et al., 2016 for details).

In comparison, developers of the Perlesko (Bizer & Karl, 2002a) used teacher ratings of test takers' vocabulary knowledge separately for each of the three language sections of the test as external variable. These ratings were correlated with children's test performances. All results are highly significant, demonstrating strong correlations between the test and the tested criteria. A similar approach was used by Haug (2011): he correlated the teachers' rating of the deaf students' DGS skills with their raw scores on the DGS Receptive Skills Test. The results showed a strong correlation. In case of the ASL-PI (Maller et al., 1999) test takers' performance scores were compared with their scores from two subtests of the Test Battery for ASL Morphology and Syntax (Supalla et al. 1995), i.e., Verbs of Motion Production and (2) Sign Order Comprehension. Bebko and colleagues (2003) used different tests to compare to test takers' performance on the Language Proficiency Profile, depending on their ages: the Expressive Communication subscale of the Vineland Adaptive Behaviour Scales (younger children: Sparrow, Balla, & Chicchetti, 1984) and the Bankson Language Screening Test (older children: Bankson, 1977) (Bebko, Calderon, & Treder, 2003) whereas Hoffmeister (2000) compared children's performance on the ASLAI with the Stanford Achievement Test (SAT-HI) and the Rhode Island Test of Language Structure (RITLS; Engen & Engen, 1983).

Predictive validity is predicted based on one or more known measured variables, such as performance on standardized achievements tests used in schools, e.g., Stanford Achievement Test (SAT-10; Pearson, 2014), Wechsler Individual Achievement Test (WIAT-II; Wechsler, 2005). For instance, the scores of the BSL Receptive Skills Test (Herman et al., 1999) were compared for children according to their years of exposure to BSL. In the youngest age groups, children from deaf families performed better than children from hearing families. For the older age groups, there was no significant difference between native singers and deaf children from hearing families on bilingual programs, however both of these achieved significantly higher scores than deaf children from hearing families on Total Communication programs. In the latter group, those children with deaf siblings or other deaf relatives achieved higher scores than those without deaf relatives. Similarly, to gather evidence for predictive validity, the developers of the DGS Receptive Skills Test used variables such as the lengths of exposure to DGS, parental hearing status, and chronological age to provide additional information that could help explaining performance differences (Haug, 2011). Evidence for predictive validity of the Assessment for NGT was gathered in form of the results by Ormel (2008), who had used the test in a previous study which showed significant correlations between children's performance on the receptive vocabulary task and their reading comprehension skills.

3.3 Reliability

Reliability refers to whether the test actually measures what it is intended to measure (e.g., Rust & Golombok, 2000). Reliability can be measured in different ways. The most commonly known ones are (1) stability over time and (2) internal consistency. The reliability of a test over time is known as test-retest reliability (Kline, 2000) for which subjects' scores

obtained on two different occasions are correlated. The higher the correlation, the more reliable is the test. The *internal consistency* of a test refers to "the degree to which scores on individual items or group of items on a test correlate with one another" (Davies et al., 1999, p. 86). A measure of internal consistency includes statistical procedures such as Cronbach's alpha. A minimum value of .70 can be considered as an "acceptable" value for a Cronbach's alpha (Nunnally, 1978).

Additional measure of reliability is inter-rater and intra-rater reliability. Inter-rater reliability refers to the level of agreement between two or more raters on a participant's performance (Davies et al., 1999), for example, by video-recording a child's language production and then comparing the scoring of specific grammatical structures by two different raters. Intra-rater reliability refers "to the extent to which a particular rater is consistent in using a proficiency scale" (Davies et al., 1999, p. 91) on different occasions. Intra-rater reliability can be established by comparing the rated scores of candidates that have been tested on two occasions that are, for example, a month apart (Davies et al., 1999) by the same raters.

As for inter-rater reliability, Hoffmeister et al. (1989) established inter-rater reliability for a narrative production test, which was part of the ASLAI by using trained raters, who evaluated the signed narratives of deaf children. Inter-rater reliability was high with .90, for both deaf and hearing raters (Hoffmeister, 1999). A similar approach was used for the Test of ASL (Strong, Prinz, & Kunze, 1994). Inter-rater reliability was investigated for each subtest "that required subjective decisions by having raters score the same set of 10 protocols, reviewing and resolving disagreements, and then scoring a second set of 10 protocols. Eventual agreement was better than 96% in all cases" (Strong & Prinz, 1997, p. 40).

Hermans et al. (2010) also established inter-rater reliability for the productive measures of the NGT Assessment Instrument. Of the five productive tasks, "13 test administers scored a randomly selected group of children within a particular age-group for the second time but now from videotape" (p. 113). Applying a Spearman's rank correlation coefficient (rho), the correlations between the raters ranged from .78 to .92, which can be considered as high.

As for intra-rater reliability, to our knowledge, there is no study that has focused specifically on intra-rater reliability.

3.4 Modern Approaches to Test Validation

In recent decades, argument-based approach has become the standard to validate (language) tests. Within this framework, validity does not include different kinds of validity (e.g., content, criterion-related and construct validity), but is rather viewed as a unified concept, i.e., construct validity (Kane, 1992). The core of validity is not to validate a test itself but the inferences made on the basis of test score interpretation and use (Messick, 1990): "test validation is empirical evaluation of the meaning and consequences of measurement" (Messick, 1990, p. 1487). Whereas reliability has been viewed as a "distinct form and a necessary *condition for validity*" (Chapelle, 1999, p. 258), validity is seen as one type of evidence for validity in more recent views on validation (Chapelle, 1999). The process of validation includes different rationales/arguments and different types of evidence that need to be investigated/collected (and used as the basis to build a validity argument).

Within an argument-based framework, five basic concepts are of importance (see also Table 2 for an example): (1) *claim* is "the conclusion of arguments that we seek to justify" (Fulcher & Davidson, 2007, p. 164), (2) *grounds* are the available evidence for the claim, (3)

warrant is the link between the evidence (grounds) and the claims, (4) *backing* is additional support for the warrant (Fulcher & Davidson, 2007), and includes, for example previous research or experience or comes from theory (e.g., Bachman, 2005). (5) The *rebuttal* is "a counter-claim that the warrant does not justify the step from the grounds to the claim." (Fulcher & Davidson, 2007, p. 165). An example for an argument-based framework for sign languages could be the argument for the German Sign Language Receptive Skills Test (Haug, 2011) for children (see Table 2).

Table 2: *Argument structure for multiple-choice items to assess the acquisition (comprehension) of morphological constructions in German Sign Language (DGS) in deaf children 4-11 years old (from Haug, 2016)*

(2) Grounds:	→	**(1) Claim:**
1. Item statistics 2. Correlation of raw scores with "external" variables such as chronological age of children, age of acquisition, and parental hearing status.		Responses to the items will allow to make claims about the acquisition process and that the parents' hearing status (deaf vs. hearing) has an influence on the DGS acquisition.
	(Since)	(Unless)
	(3) Warrant:	**(5) Rebuttal:**
	Sign language acquisition research has shown that	1. Younger children achieve higher raw

	certain morphological constructions (across sign languages) are only acquired (mastered) when children are 10-12 years old (compared to constructions that are mastered by age 6) and that deaf children of deaf native signing parents acquire a sign language as their first language (compared to deaf children of hearing parents who might learn a sign language later). **(4) Backing**: Review of research studies that focuses at morphological constructions in DGS and the acquisition (from emergence to mastery) of these constructions in deaf children.	scores than older children 2. Deaf children of hearing parents achieve higher raw scores than deaf children of deaf parents 3. Test items do not represent DGS constructions

However, alternative types of statistical methods are setting new standards within language testing to investigate reliability, for example, multi-facet Rasch measurement (Linacre, 1994) can investigate the interactions between different facets, such the items and the raters. Rasch measurement "is an attempt to model the relationship between various facets of the test situation" (McNamara, 1996, p. 154).

4. Pedagogical Applications

This next section offers an overview of, tests from different sign languages that are used by practitioners within the school context. It is arranged based the following criteria: (1) assessment target, e.g., vocabulary, grammar, (2) type of assessment: receptive and/or productive skills test, (3) target group, e.g., babies, toddlers, children, (4) language for which the test was originally developed, and (5) sign language(s) for which the test has been adapted. We acknowledge that there may be other sign language tests, which are not mentioned in this chapter. The reasons for not including these tests is that they may not have been published on (in a language that is accessible to the authors), focus on a different target population, e.g., adults, and/or are not available to practitioners.

4.1 Assessing Vocabulary in Deaf Children

4.1.1 The MacArthur-Bates Communicative Development Inventory (CDI)

Assessment target: Language development in monolingual hearing children. The CDI examines comprehension, word production, and early phases of grammar.

Format: Standardized parental checklist. Parents complete the checklist in regular time intervals by ticking off any words or signs that their child can understand and/or is able to produce.

Target population: Hearing children between 8-36 months

Developed for: American English (CDI, Fenson et al., 1994)

Adapted for: American Sign Language (ASL; Anderson & Reilly, 2002), British Sign Language (BSL; Woolfe, Herman, Roy, & Woll, 2010).

4.1.2 Perlesko: Vocabulary Test for German Sign Language

Assessment target: Language development in signing children, specifically receptive vocabulary.

Format: The Perlesko uses a multiple-choice format, which requires children to either match a signed or spoken word to one of four picture choices or match a picture to one of four words in written German. In addition to German Sign Language (DGS), it can also be used to assess children's comprehension skills in spoken German, and written German.

Target population: Deaf children between 7-13 years

Developed for: German Sign Language (Bizer & Karl, 2002b)

Adapted for: N/A

4.1.3 British Sign Language Vocabulary Test

Assessment target: Language development in signing children, specifically receptive and productive vocabulary

Format: The BSL-VT is a web-based instrument, which measures strength of vocabulary knowledge. The test consists of four tasks, i.e., form recall (production), form recognition (comprehension), meaning recall (production), meaning recognition (comprehension), each of which assesses a different level of deaf children's vocabulary knowledge. The same items are

used across all tasks to provide test administrators with a more detailed measurement of children's knowledge of each sign in an effort to guide and improve intervention (Mann, Roy, & Marshall, 2013).

Target population: Deaf children between 4-15 years

Developed for: BSL (Mann & Marshall, 2012)

Adapted for: ASL (Mann, Roy, & Morgan, 2016), Finish Sign Language (Kanto & Mann, in preparation).

4.2 Assessing grammatical aspects in deaf children

4.2.1 British Sign Language Receptive Skills Test

Assessment target: Language development, specifically receptive knowledge of the following BSL syntactic and morphological structures: (1) spatial verb morphology, (2) number and distribution, (3) negation, (4) size/shape specifiers, (5) noun-verb distinction, and (6) handling classifiers.

Format: Vocabulary check and a video-based receptive skills test. Prior to the receptive skills test, a vocabulary check is conducted. The children confirm their knowledge of the 22-item vocabulary used in the main test through a simple picture-naming task that identifies signs taken from the receptive skills test. There are two versions of this task, one for the North and one for the South of the UK

Target population: Deaf children between 3-11 years.

Developed for: BSL (Herman et al., 1999; Herman, Rowley, & Woll, 2015).

Adapted for: ASL (Enns & Herman, 2011) and German Sign Language (Haug, 2011), Finnish Sign Language (Kanto, in progress), Polish Sign Language (Enns et al., 2016), and Spanish Sign Language (Valmaseda, Perez, Herman, Ramírez, & Montero, 2013)

4.2.2. ASL Assessment Instrument

Assessment target: Conversational abilities, academic language knowledge, language comprehension, analogical reasoning, and metalinguistic skills

Format: The ASLAI is a web-based test consisting of 12 tasks. Items are presented in multiple-choice format. Tasks in the ASLAI are in one of six formats: 1) picture to sign, 2) sign to sign, 3) picture to picture, 4) drag-and-drop sorting, 5) response-only (grammaticality judgment), and 6) video event to sign

Target population: Deaf children between 4-18 years

Developed for: ASL (Hoffmeister, Caldwell-Harris, Henner, Benedict, Fish, Rosenburg, Conlin-Luippold, & Novogrodsky, 2014)

Adapted for: N/A

4.2.3 Assessment Instrument for Sign Language of the Netherlands

Assessment target: Phonology, morpho-syntax, and narrative skills (both receptive and productive)

Format: This assessment instrument is a computer-based test consisting of nine tasks. Formats include multiple-choice (e.g., receptive morpho-syntactic, task), or retelling a picture story shown on screen.

Target population: Deaf children between 4-12 years

Developed for: Sign Language of the Netherlands (NGT, Hermans et al., 2010)

Adapted for: N/A

4.2.4 BSL Productive Skills Test

Assessment target: Narrative skills and BSL grammar

Format: The BSL PST is a computer-based test. It uses a narrative recall format in which children watch a short language-free video and retell the story and answer questions targeting comprehension and inferencing skills.

Target population: Deaf children between 4-11 years.

Developed for: BSL

Adapted for: ASL (Enns, Boudreault, Zimmer, Broszeit & Goertzen, 2014), Australian Sign Language (Hodge, Schembri & Rogers, 2014), Spanish Sign Language (Enns et al., 2016), and English (Jones, Herman, Botting, Marshall, Toscano, & Morgan, 2015)

4.2.5 Nonsense Sign Repetition Task (NSRT)

Assessment target: Phonological development in sign language

Format: The NSRT is a web-based test in which participants repeat pre-recorded, nonsense signs of differing phonetic (i.e., handshape and movement) complexity.

Target population: Deaf children between 3-11 years

Developed for: BSL (Mann, Marshall, Mason, & Morgan, 2010)

Adapted for: Icelandic Sign Language (Ivanova, 2012)

4.2.6 Test of ASL (TASL)

Assessment target: Morpho-syntactic skills (both receptive and productive)

Format: The TASL is a video-based test consisting of six tasks. Formats include multiple-choice (e.g., classifier comprehension, map marker task) or retelling a picture story from a book without text.

Target population: Deaf children between 8-15 years

Developed for: ASL (Strong, Prinz, & Kuntze, 1994).

Adapted for: Swedish Sign Language (Schönström, Simper-Allen, & Svartholm, 2003); French Sign Language used in Switzerland (Prinz, Niederberger, Gargani, & Mann, 2005)

4.2.7. Instrumento de avaliação da língua de sinais brasileira (IALS)

Assessment Target: Morpho-syntactic skills (both receptive and productive)

Format: In the comprehension task here are a set of pictures for each level evaluated combined with a video in Libras telling a short story. The stories are produced involving different levels of vocabulary, uses of space and adding referents. The participant must to watch the video and choose the pictures related to what was signed ordering the pictures considering the story told. The stories increase the complexity of matching the levels of the language developed. The second part involves production. The participant watches a short story in the video, then tells the story to someone else. This signing production is evaluated considering a chart with criteria considering the number of the events, vocabulary, uses of space, classifiers and general vocabulary.

Target population: children from 4 to 9 years old and late learners

Developed for: Libras (Brazilian Sign Language) (Quadros & Cruz, 2011)

Adapted for: NA

5. Use of Sign Language Assessments in Practice

Whilst many assessments for signed languages have been developed over the course of the past decade, their use in everyday education and health settings raises some issues. These include accessibility, purpose, training, and intervention planning. Unfortunately, there is very limited literature on practitioners' use of (sign) language assessments. In order to encourage awareness and use of the assessments in enabling children to learn language as effectively as possible, these issues need to be addressed.

Information about language assessment tools is (more) readily available in academic papers and at conferences. However, these are not always the most effective means of informing practitioners working in schools and clinics. Even if the teachers, therapists and assistants access this information, it is not always easy to translate a research tool described in academic terms into a functional procedure for use in the classroom (Hoskin, 2017). In the UK, the Deafness Cognition and Language Research Centre (DCAL) has tried to address some of these concerns by making assessment tools available via a website (www.dcalportal.org). Specifically, this has enabled practitioners to more easily access assessments that are relevant to their work with signing deaf individuals. Additionally the consistent use of these instruments by practitioners enables researchers/test developers to periodically review the norms for their tests, e.g., BSL Receptive Skills Assessment (Herman, Homes, & Woll, 1999). However, whilst online accessibility has made this and other tools readily available for clinical use, it raises the need for local protocols to be written as practitioners are asked to enter confidential details e.g. name, date of birth, and schools and clinics need to ensure their staff are not breaking confidentiality rules.

This process is not without problems as illustrated in the following comment by a UK-based practitioner:

"*The confidentially issue has been the main barrier to me using the test. I understand that the language of the confidentiality agreement needs to be formal and detailed but I find that 'off putting', and sometimes inaccessible, to parents/guardians. I'd like something simple along the lines of: 'I'm using this new, interesting assessment with your child in school. The researchers will be using the scores and ages of the children, anonymously to improve and develop much needed assessments. Is that ok?"*

*But it's always much more complicated than that and I understand why from the university point of view. Also the *** test website asks for quite a lot of demographic data. With older children who have for example come to the school in Year 9, this demographic detail isn't shown in the records and I would need to ask parents/guardians so it becomes a bigger job for the person administering the test and potentially reduces parent/guardian confidence that letting the test results be used by researchers is no big deal. It becomes a bigger more complicated issue and I end up having this kind of sinking feeling about the website although I find the assessments exciting and want to get on and use them*." (Dumbrill, personal communication)

As more assessments for spoken language become accessible online, for example, the Clinical Evaluation and Language Fundamentals (CELF-5; Wiig, Semel, & Secord, 2013) and practitioners become accustomed to using online tools, some of the issues discussed above may be resolved. As practitioners grow more accustomed to incorporate online tools in their assessment routine, this is likely to lead to be a broader understanding of how to use the online format and how to deal with issues around confidentiality. In turn, this will raise manager and

supervisor awareness of issues related to training, funding and supervision linked to the use of online assessment tools.

Another issue that may affect practitioners' use of sign language assessments is that many of the existing sign language tests have been developed as part of research projects. As a consequence they tend to measure aspects of language for a purpose linked to a very specific research question. Understanding how and when to use these tests with children developing signed languages often requires at least a certain knowledge of language development and disorder. For the researchers using the tests within projects, background reading and training on test-use is part of the project protocol, yet for practitioners it is not. Thus explicit instructions for completing a test, and opportunities for practitioners to practice test use are needed to ensure they are confident with materials and procedures before trying to use a tool with a child. Yet, this theoretical learning and practical training is not always available to practitioners. In some educational settings, the assessment tools that have been developed are used regularly to monitor children's progress, in others the tools are used when there are concerns about a child's language learning abilities. There are also settings where practitioners report that they do not use them, as although they have access to the tools, they are not sure how to interpret the results or how the results would be useful in developing any intervention. Where tools have been developed that link to interventions regularly used, the papers reporting them require the practitioner to be proactive in seeking detail that would enable their use with a child. One example of this is a paper on 'narragrams' where cartoon stories from a tv show are broken into mini-events for assessing children's narrative skills (Erber, Grant, Leigh, & Kenfield, 2016). However, the paper does not contain the story titles or mini event lists, requiring the practitioner to seek these from the authors. While this is not a difficult task, it is an additional one which may reduce the clinical

use of excellent tools. Providing this information may not be possible or appropriate for authors, depending on publication criteria, however, a facility that linked such papers and practical research outcomes with the accessible product for the practitioner to use would be of considerable benefit to both researchers and practitioners. One facility currently exists that brings together information about sign language tests, the use of the tests and the research background to their development (http://www.signlang-assessment.info/index.php/home-en.html). It highlights that many tests are not commercially available, suggesting people interested contact the developers direct. Whilst this is a useful website for researchers and academics, the test accessibility and usability to a practitioner in a clinic or educational setting is limited.

For many practitioners one key role of tests is to guide intervention planning. However, without appropriate knowledge of language development and disorder, intervention planning can become 'teaching a test' to a child, which must be avoided. Consideration of the workforce is needed as some practitioners working with children who sign do not have native levels of the target language themselves and therefore may have difficulty in both assessing accurately and translating the results into intervention targets and strategies. Within the workforce there will be both deaf and hearing practitioners with variability around whether individuals are native signers or the level to which they understand and use sign language. This will not only affect test-use but also the practitioners' access to training courses regarding assessments unless these are provided in their strongest and preferred language.

As described above, researchers have worked hard to develop sign language assessments and practitioners are striving to improve their assessment and intervention of children's signed language by using these tools. However, there are still challenges that need to be overcome.

Accessibility, confidentiality, and training are some key issues that need attention to support the work of practitioners and provide benefits for children. These issues need to be addressed to enable a more effective use of current assessments. One way to do this could be in form of online resources (webinars) to demonstrate test use similar to those offered for many spoken language assessments, e.g., the CELF-5. Another way forward could be researchers and clinicians working together to translate a research tool, i.e., test, into a functional procedure for use in the classroom. To make such collaboration successful researchers and practitioners need time to work together to really understand the concept of joint working in this context, to remove barriers and to ensure easy and productive communication around the practice. This could be approached through joint working groups or focus groups. Practitioners also have service managers, parent/guardians of clients and clients themselves whom they need to take with them on this venture. Researchers making their assessments available online, practitioners using them and feeding back data is not yet a concept that *routinely* exists. However, if the concept and process were properly established and embedded in the everyday working life of researchers and practitioners, that would be exciting and beneficial.

6. Technology and Sign Language Testing

One aspect of sign language testing that is considered highly technological is test delivery. The use of computer- or web-based test formats for delivery are obvious advantages for sign language tests (e.g., Haug, 2015). Web-based tests are particularly useful as they allow automatic scoring of multiple-choice tests and are easily accessible from anywhere in the world (given that high-speed Internet is available) (Haug, 2015).

The shift from traditional paper and pencil format of testing towards web-based formats for test delivery has reached the field of sign language assessment, as demonstrated, for example, by the BSL Vocabulary Test (Mann & Marshall, 2012) and the BSL Receptive Skills Test (Herman et al., 1999). In addition, these tests are part of a web-based assessment portal set up by the Deafness, Cognition and Research Centre at University College London. Another example of using a web-based format for sign language assessment is the narrative comprehension test for Swiss German Sign Language (*Deutschschweizerische Gebärdensprache,* DSGS; Haug & Perrollaz, 2015) which has been developed within the frame of the EU project SignMET. Similar to the BSL-VT and BSL-RST, this test is integrated in a purpose-build portal for sign language tests[2].

One of the biggest disadvantages of web-based tests, according to Haug (2015), are difficulties with the technical infrastructure (e.g., old hardware and software, server connectivity). Despite the assumption that most test takers should be familiar with the use of a computer or mobile device, this may not always be the case so that a test taker's level of computer familiarity might have an impact on the test results (for a detailed discussion on web- and mobile-based testing formats, see Haug, 2015).

There are other technologies that show potential use for sign language assessment. One example is automatic sign language recognition. The focus of a recent Swiss National Science Foundation project SMILE[3] (Ebling et al., 2018) makes use of this technology. One of the goals of the SMILE project is to develop an automatic sign language recognition system that will be

[2]https://signlang-portal.com

[3] The SMILE project is a consortium of three institutions: the Idiap Research Institute in Martigny, Switzerland (lead), the University of Applied Sciences for Special Needs Education in Zurich (Switzerland), and the University of Surrey (United Kingdom).

used in the context of vocabulary assessment for adult L2 learners of DSGS. The set-up for such a testing scenario is that a test taker will be asked to produce a lexical sign (e.g., with a gloss or translating a written German word into DSGS), delivered on a computer screen. The produced sign will be recognized by a camera and compares the test taker's performance with the "correct" or "acceptable" form of the sign the recognition system has been trained and provide feedback to the learner if the sign is produced correctly or not.

7. Future Trends

7.1 Future research studies

Apart from the fact that basic research on the structure and acquisition of sign languages is needed, we will focus here on two different strands of research. The first strand on future research studies is the development and evaluation of rating scales of productive sign language tests, the second strand are studies on the use of new technologies in sign language testing.

Some of presented sign language tests include rating scales of production (for example, Strong & Prinz, 1997). Issues like inter-rater reliability have been investigated, but in the future we should also try to focus on issues if raters have a mutual understanding of the rating criteria, if the underlying construct of the rating scale if clearly understood, how do raters solve possible disagreement between their ratings (e.g., Haug, Batty, & Ebling, in preparation). Investigating such issued can contribute to the validity of the rating scale.

As for the second strand, the use of new technologies, possible future studies could focus on sign language technologies, such as sign language recognition (SLR) and how they could be applied in the domain of sign language assessment. One example is the above mentioned project SMILE which uses SLR to score test taker' performance in a vocabulary test.

7.2 Future Pedagogical Applications

In the beginning of this chapter, we have mentioned the few sign language tests for children that are commercially available. The need for such tests in schools has been pointed out in different studies (e.g., Haug & Hintermair, 2003). An area is the provision of training for (future) teachers of the deaf, SLT/SLPs, Deaf practitioners, and sign language tutors on how to use these tests both with regard to administration as well as score interpretation and what this means for support. Finally, despite their value for professionals working with deaf individuals most existing (sign) language tests have limitations in that the information they provide focuses on the child's learning outcome rather than the learning process. This calls for alternative methods of testing, for instance dynamic assessment, which enables practitioners to make assumptions about children's response to a particular type of intervention in the future. The research done in this area, while limited (Mann, 2017; Mann, Peña, & Morgan, 2014, 2015), looks promising.

References

Anderson, D., & Reilly, J. (2002). The MacArthur communicative development inventory: normative data for American Sign Language. *Journal of Deaf Studies and Deaf Education*, *7*(2), 83–106. https://doi.org/10.1093/deafed/7.2.83

Bankson, N. W. (1977). *Bankson language screening test*.

Bebko, J. M., & McKinnon, E. E. (1993). *The Language Proficiency Profile-2*. Unpublished assessment tool. York University, Toronto, Canada.

Bebko, J. M., Calderon, R., & Treder, R. (2003). The Language Proficiency Profile-2: Assessment of the global communication skills of deaf children across languages and modalities of expression. *Journal of Deaf Studies and Deaf Education*, *8*(4), 438–451. https://doi.org/10.1093/deafed/eng034

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, *2*(1), 1–34. https://doi.org/10.1207/s15434311laq0201_1

Bechtoldt, H. P. (1951). Selection. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1237–1267). New York, NY: Wiley.

Bizer, S., & Karl, A.-K. (2002a). *Entwicklung eines Wortschatztests für gehörlose Kinder im Grundschulalter in Gebärden-, Schrift- und Lautsprache*. Unpublished doctoral dissertation, Fachbereich Erziehungswissenschaften, Universität Hamburg.

Bizer, S., & Karl, A.-K. (2002b). *Perlesko: Prüfverfahren zur Erfassung lexikalisch-semantischer Kompetenz gehörloser Kinder im Grundschulalter*.

Boyes Braem, P., Haug, T., & Shores, P. (2012). Gebärdenspracharbeit in der Schweiz: Rückblick und Ausblick. *Das Zeichen*, *90*, 58–74.

Bush, C. T. (1985). *Nursing research*. Virginia: Reston Publishing Company.

Chapelle, C. A. (1999). Validity in language assessment. *Annual Review of Applied Linguistics*, *19*, 254–272.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, *52*(4), 281–302.

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of language testing – Studies in language testing 7*. Cambridge: Cambridge University Press.

Dumbrill, H. (2018). Personal communication, September 2nd, 2018.

Ebling, S., Camgöz, N. C., Boyes Braem, P., Tissi, K., Sidler-Miserez, S., Stoll, S., … Magimai-Doss, M. (2018). SMILE Swiss German Sign Language data set. *11th Language Resources and Evaluation Conference (LREC 2018)*, 4221–4229.

Engen, E., & Engen, T. (1983). *Rhode Island test of language structure*. Pro-Ed, Incorporated.

Enns, C., Boudreault, P., Zimmer, K., Broszeit, C., & Goertz, D. (2014). *Assessing children's expressive skills in American Sign Language*. Paper presented at the Annual Meeting of the Association of College Educators – Deaf/Hard of Hearing. Washington, DC.

Enns, C., & Herman, R. (2011). Adapting the assessing British Sign Language development: Receptive Skills Test into American Sign Language. *Journal of Deaf Studies and Deaf Education, 16*(3), 362–374. https://doi.org/10.1093/Deafed/enr004.

Enns, C., Haug, T., Herman, R., Hoffmeister, R. J., Mann, W., & Mcquarrie, L. (2016). Exploring signed language assessment tools in Europe and North America. In M. Marschark, V. Lampropoulou, & E. K. Skordilis (Eds.), *Diversity in deaf education* (pp. 171–218). Oxford; New York: Oxford University Press.

Erber, N. P., Grant, L. M., Leigh, K., & Kenfield, S. (2016). The narragram: A new way to describe children's recall of stories. *Deafness & Education International*, *18*(3), 120–123. https://doi.org/10.1080/14643154.2016.1169617

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., ... & Stiles, J. (1994). Variability in early communicative development. *Monographs of the society for research in child development*, i-185.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. London; New York: Routledge.

Haug, T. (2005). Review of sign language assessment instruments. *Sign Language & Linguistics*, *8*(1/2), 61–98. https://doi.org/10.1075/sll.8.1.04hau

Haug, T. (2011). *Adaptation and evaluation of a German Sign Language test - A computer-based receptive skills test for deaf children ages 4–8 years old*. Hamburg: Hamburg University Press. Retrieved from http://hup.sub.uni-hamburg.de/purl/HamburgUP_Haug_Adaption

Haug, T. (2015). Use of information and communication technologies in sign language test development: Results of an international survey. *Deafness & Education International*, *17*(1), 33–48. https://doi.org/10.1179/1557069X14Y.0000000041

Haug, T. (2016). *Example of an argument structure for multiple-choice items to assess the acquisition (comprehension) of morphological constructions in German Sign Language (DGS) in deaf children 4-11 years old.* Unpublished material.

Haug, T. (2017). *Web-based sign language assessment: Challenges and innovations*. Paper presented at the ALTE 6th International Conference: Learning and Assessment: Making the Connection, Bologna.

Haug, T., & Hintermair, M. (2003). Ermittlung des Bedarfs von Gebärdensprachtests für gehörlose Kinder - Ergebnisse einer Pilotstudie. *Das Zeichen*, *64*, 220–229.

Haug, T., Mann, W., Boers-Visker, E., Contreras, J., Enns, C., & Rowley, K. (2016). *Guidelines for sign language test development, evaluation, and use*. Retrieved on July 25 2017 http://www.signlang-assessment.info/tl_files/signlanguage/guidelines_sign_language_tests_haugetal_v1_2016-11-10.pdf

Haug, T., & Perrollaz, R. (2016). *EU-Projekt SignMET: Gebärdensprachtests im Test. Sonos*, *3*, 8–10.

Herman, R. (2002). *Assessment of BSL development*. Unpublished doctoral dissertation, City University London.

Herman, R., Holmes, S., & Woll, B. (1998). *Design and standardisation of an assessment of British Sign Language development for use with young deaf children: Final report, 1998*. Unpublished manuscript, City University London, UK.

Herman, R., Holmes, S., & Woll, B. (1999). *Assessing BSL Development - Receptive Skills Test*. Coleford, UK: The Forest Bookshop.

Herman, R., Rowley, K. & Woll, B. (2015) BSL Receptive Skills Test (RST) online. https://dcalportal.org/tests

Hermans, D., Knoors, H., & Verhoeven, L. (2010). Assessment of sign language development: The case of deaf children in the Netherlands. *Journal of Deaf Studies and Deaf Education, 15*(2), 107–119. https://doi.org/10.1093/deafed/enp030.

Hodge, G., Schembri, A. & Rogers, I. (2014) *The Auslan (Australian Sign Language) Production Skills Test: Responding to challenges in the assessment of deaf children's signed*

language proficiency. Paper presented at Disability Studies in Education Conference, July, Melbourne, Australia.

Hoffmeister, R. J. (1999). *American Sign Language Assessment Instrument (ASLAI)*. Manuscript, Center for the Study of Communication & the Deaf, Boston University.

Hoffmeister, R. J. (2000). A piece of the puzzle: ASL and reading comprehension in deaf children. In C. Chamberlain, J. P. Morford & R. Mayberry (Eds.), *Language acquisition by eye* (pp. 143–163). Mahwah, NJ: Lawrence Erlbaum Publishers.

Hoffmeister, R. J., Caldwell-Harris, C.L., Henner, J., Benedict, R., Fish, S., Rosenburg, P. Conlin-Luippold, F. & Novogrodsky, R. (2014). *The American Sign Language Assessment Instrument (ASLAI): Progress report and preliminary findings*. Working paper: Center for the Study of Communication and the Deaf.

Hoffmeister, R., Greenwald, J., Bahan, B., & Cole, J. (1989). *The American Sign Language Assessment Instrument*. Unpublished instrument. Boston University Center for the Study of Communication and the Deaf.

Hoskin, J. (2017) *Language therapy in British Sign Language: A study exploring the use of therapeutic strategies and resources by deaf adults working with young people who have language learning difficulties in British Sign Language (BSL)*. Doctoral thesis (Ph.D), UCL (University Colleg London). http://discovery.ucl.ac.uk/10022827/

Ivanova, N. (2012). *Nonsense repetition test for ÍTM*. The Communication Centre for the Deaf and Hard of Hearing: Reykjavik.

Jones, A., Herman, R., Botting, N., Marshall, C., Toscano, E., & Morgan, G. (2015). *Narrative skills in deaf children: Assessing signed and spoken modalities with the same te*st. Paper presented at the SRCD preconference: Development of deaf children. Philadelphia, PA.

Kane, M. (1992). An argument-based approach to validity. *Psychological Bulletin*, *112*(3), 527–535.

Kanto, L. (in progress): *Finish Sign Language receptive skills test.*

Kanto, L. & Mann, W. (in preparation). *Finnish Sign Language vocabulary test.*

Kline, P. (2000). *Handbook of psychological testing* (2nd. ed.). London & New York: Routledge.

Linacre, J. M. (1994). *Many-facet Rasch measurement*. Chicago, IL: MESA Press.

Maller, S., Singleton, J. L., Supalla, S. J., & Wix, T. (1999). The development and psychometric properties of the American Sign Language proficiency assessment (ASL-PA). *Journal of Deaf Studies and Deaf Education*, *4*(4), 249–269. https://doi.org/10.1093/deafed/4.4.249

Mann, W. (2017). Measuring deaf learners' language progress in school. In M. Marschark & H. Knoors (Eds.), *Evidence-based practice in deaf education*. Oxford; New York: Oxford University Press.

Mann, W., & Haug, T. (2014). Mapping out guidelines for the development and use of sign language assessment: Some critical issues, comments and suggestions. In D. Quinto-Pozos (Ed.), *Mulitilingual aspects of signed language communication and disorders* (pp. 123–139). Bristol, UK: Multilingual Matters.

Mann, W., Peña, E. D., & Morgan, G. (2015). Child modifiability as a predictor of language abilities in deaf children who use American Sign Language. *American Journal of Speech-Language Pathology*, *24*(3), 374–385. https://doi.org/10.1044/2015_AJSLP-14-0072

Mann, W., Peña, E. D., & Morgan, G. (2014). Exploring the use of dynamic language assessment with deaf children, who use American Sign Language: Two case studies. *Journal of Communication Disorders*, *52*, 16–30. https://10.1016/j.jcomdis.2014.05.002

Mann, W., & Marshall, C. (2012). Investigating deaf children's vocabulary knowledge in British Sign Language. *Language Learning*, *62*(4), 1024–1051. https://doi.org/ 10.1111/j.1467-9922.2011.00670.x

Mann, W., Marshall, C. R., Mason, K., & Morgan, G. (2010). The acquisition of sign language: The impact of phonetic complexity on phonology. *Language Learning and Development*, *6*(1), 60–86. https://doi.org/10.1080/15475440903245951

Mann, W., Roy, P., & Marshall, C. (2013). A look at the other 90 per cent: Investigating British Sign Language vocabulary knowledge in deaf children from different language learning backgrounds. *Deafness & Education International*, *15*(2), 91–116. https://doi.org/ 10.1179/1557069X12Y.0000000017

Mann, W., Roy, P., & Morgan, G. (2016). Adaptation of a vocabulary test from British Sign Language to American Sign Language. *Language Testing*, *33*(1), 3–22. https://doi.org/10.1177/0265532215575627

Mayberry, R. I., Lock, E., & Kazmi, H. (2002). Development: Linguistic ability and early language exposure. *Nature, 417*, 38. http://dx.doi.org/10.1038/417038a

McNamara, T. (1996). *Measuring second language performance*. London: Longman.

Messick, S. (1990). Validity of test interpretation and use. *ETS Research Report Series*, *1990*(1), 1487–1495. https://doi.org/10.1002/j.2333-8504.1990.tb01343.x

Newport, E. L. (2002). Critical periods in language development. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (pp. 737–740). London: Macmillan Publishers Ltd.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Ormel, E. A. (2008). *Visual word recognition in bilingual deaf children*. Doctoral dissertation. Radboud University Nijmegen.

Palfreyman, N., Sagara, K., & Zeshan, U. (2015). Methods in carrying out language typological research. In E. Orfanidou, B. Woll, & G. Morgan (Eds.), *Research methods in sign language studies: A practical guide* (pp. 173–192). Chichester, UK: Wiley Blackwell.

Pearson (2014). *Stanford Achievement Test. 10th ed. (SAT-10).*

Prinz, P. M., Niederberger, N., Gargani, J., & Mann, W. (2005, July). *Cross-linguistic and cross-cultural issues in the development of tests of sign language: The case of the test of American Sign Language (TASL) and Test de Langue des Signes Francaise (TELSF)*. Paper presented at the IASCL, Berlin.

Quadros, R. M. de, & Cruz, C. R. (2011) *Instrumento de avaliação: língua de sinais*. Editora ArtMed. Porto Alegre.

Rust, J., & Golombok, S. (2000). *Modern psychometrics – The science of psychological assessment* (2nd ed.). London & New York: Routledge.

Schönström, K., Simper-Allen, P., & Svartholm, K. (2003). Assessment of signing skills in school-aged deaf students in Sweden. In *European Days of Deaf Education, May 8-11, 2003, Örebro, Sweden* (pp. 88–95). Örebro, Sweden: Stockhom University.

Sparrow, S. S., Balla, D. A., Cicchetti, D. V., Harrison, P. L., & Doll, E. A. (1984). *Vineland adaptive behavior scales*.

Strong, M., Prinz, P., & Kuntze, M. (1994). *The Test of ASL*. Unpublished test. San Francisco: San Francisco State University, California Research Institute.

Strong, M., & Prinz, P. (1997). A study of the relationship between American Sign Language and English literacy. *Journal of Deaf Studies and Deaf Education, 2*(1), 37–46.

Supalla, T., Newport, E., Singleton, J. L., Supalla, S. J., Metlay, D., & Coulter, G. (1995). *An overview of the Test Battery for American Sign Language Morphology and Syntax*. Paper

presented at the Annual Meeting of the American Educational Research Association (AERA), San Francisco, CA.

Valmaseda, M., Pérez, M., Herman, R., Ramírez, N & Montero, I. (2013) *Evaluación dla competencia gramatical en LSE: Proceso de adaptación del BSL Receptive Skill Test (test de habilidades receptivas)*. Retrieved on July 30 2018 from www.cnlse.es/sites/default/files/Evaluacion%20de%20la%20competencia%20gramatical%20en%20LSE.pdf.

Wiig, E., Semel, E., & Secord, W. A. (2013). *Clinical evaluation and language fundamentals (5th ed.) (CELF-5)*. Bloomington, MD: Pearson Clinical.

Van Herreweghe, M., & Vermeerbergen, M. (2009). Flemish Sign Language standardisation. *Current issues in language planning*, *10*(3), 308–326. https://doi.org/10.1080/14664200903154874

Wechsler Individual Achievement Test (WIAT II) (2005). London: The Psychological Corp.

Woll, B. (2013). Second language acquisition of sign language. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Oxford, UK: Blackwell Publishing Ltd. http://doi.wiley.com/10.1002/9781405198431.wbeal105

Woolfe, T., Herman, R., Roy, P., & Woll, B. (2010). Early vocabulary development in deaf native signers: A British Sign Language adaptation of the communicative development inventories. *Journal of Child Psychology and Psychiatry*, *51*(3), 322–331. https://doi.org/10.1111/j.1469-7610.2009.02151.x

Yaghmaei, F. (2003). Content validity and its estimation, *Journal of Medical Education*, 3(1), 25–27.

Useful links:

Overview if sign language assessments

http://www.signlang-assessment.info/index.php/home-en.html

Guidelines for developing sign language assessment

https://www.hfh.ch/fileadmin/files/documents/Zentrale_Dienste_Personal/who_is_who/hat_guidelines_sign_language_tests_haugetal_v1_2016-11-10.pdf

DCAL Assessment Portal

http://www.ucl.ac.uk/dcal/dcal-portal

Sign Language Portal of Sign Language Assessment Instruments

https://signlang-portal.com